

Bergé, Laurent

Conference Paper

Closure in inter-regional knowledge networks: An application to the European co-publication network

54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Bergé, Laurent (2014) : Closure in inter-regional knowledge networks: An application to the European co-publication network, 54th Congress of the European Regional Science Association: "Regional development & globalisation: Best practices", 26-29 August 2014, St. Petersburg, Russia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/124507>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Closure in inter-regional knowledge networks: An application to the European co-publication network

Laurent R. Bergé*

July 26, 2014

Abstract

The question of the determinants of cross-regional knowledge flows has received a growing interest in the recent past. Particularly, the question of the relationship between geography and networks has been debated. This paper contributes to this debate by investigating the effect of network-closure on cross regional scientific collaborations. The concept of closure relates to fact that indirect connections in networks facilitates direct connections. Contrary to other studies focusing at the individual (micro) level, we introduce a methodology that allows to work directly at the regional level. This change of scale avoids problems rising at the micro level such as the identification of individuals and also allows much larger scale studies. The network closure between regions is here captured by an index based on a simple model of random matching between agents. We then assess the effect of network closure by using data on scientific collaborations in the field of Chemistry in Europe. The results show a significant effect of the index of closure in determining cross-regional knowledge flows. It also significantly alleviates the impeding effect of national borders on cross-countries collaborations, then suggesting that network closure may be a channel favored for international collaborations.

Keywords: network formation, gravity model, regional closure, aggregated networks, spatial proximity, network proximity, co-publication; research collaboration

JEL codes: D85, O31, R12

1 Introduction

The production of knowledge is a central issue for policy makers, at the regional, national or even supra-national scale. Indeed, knowledge is viewed as essential in enhancing competitiveness and producing long-term growth ([Aghion and Howitt, 1992](#)). And, as stated in the

*University of Bordeaux. GREThA – UMR CNRS 5113

'Lisbon agenda', the European Union (EU) is particularly concerned by being a world leading knowledge-based economy. Also, along policies targeting an increased level of research and development from its member states, the EU aims at creating an European research area (ERA) where "national systems must be more open to each other and to the world, more inter-connected and more inter-operable" (European commission, 2012, p. 3). This pursuit of an increased level of collaboration and coordination at all geographical levels, comes along with the fact that knowledge production becomes increasingly reliant on collaboration (e.g. Wuchty et al., 2007; Jones et al., 2008). As more resources are needed in the collaborative knowledge production process, their good allocation is critical to have an efficient ERA. And as a corollary, the barriers hindering collaboration should be as low as possible to allow the best matches to occur. This notion of ERA intermingles at the same time a spatial characteristic, with a focus on EU geographical areas and integration, and an a-spatial one, with the need of connecting the actors in the most efficient way. To assess whether the ERA is reaching its goal, there is a need to understand the interplay between a-spatial networks and geography (Maggioni and Uberti, 2011).

The purpose of this paper is to capture to what extent the knowledge flows occurring between regions are dependent on the structural pattern of the inter-regional network. To this end, we first introduce a measure specifically designed for aggregated network to capturing the strength of the indirect connections at play between the dyads of regions; and second, we assess the effect of this endogenous network effect and its relation with other determinants of inter-regional knowledge flows in the context of the European co-publication network. The main idea of the paper is that relational ties across regions carries informational externalities at the regional level. We argue here that the concept, first defined for inter-individual relations, of triadic closure also holds for the inter-regional network of knowledge flows, though slightly modified. This concept is proxied by using the idea of inter-regional bridging path. Here a bridging path is defined as an indirect connection between two regions via a third 'bridging region' at the micro-level. Then from a simple random matching process, modeling the collaborations among the micro-level actors based on the information of the aggregated level, we derive a closed form of the expected number of bridging paths between two regions. This measure is then used to assess 'regional closure', that is, the propensity of two regions to have a link with respect to their indirect connections. This relates to the notion of some kind of 'network proximity', reflecting the idea that the network is a purveyor of externalities.

This study is then in line with a recent stream in the literature analyzing the determinants of knowledge flows across regions (e.g. LeSage et al., 2007; Hoekman et al., 2009; Frenken et al., 2009). Yet, best to our knowledge, no study has specifically investigated the role of closure, or network related effect, in inter-regional knowledge networks. This fact may come from mainly two reasons. First, as the inter-regional network of knowledge flows is an aggregated network of individual decisions, the tools to assess the extent to which two

regions are indirectly connected may be missing. Second, it is conceptually difficult to treat regions as entities behaving as individuals and being subject to the same concepts, as the one of closure. Our methodology allows to challenge these two potential issues.

To test empirically whether the network closure has an effect on the inter-regional knowledge flows, we make use of European co-publication data. More specifically, we retrieve information on scientific publications from the Thomson Reuters Web of Science database. It consists of the co-publications stemming from the five largest European countries (France, Germany, Italy, Spain, the United Kingdom), from journals of chemistry for the period 2001-2005. The analysis consists on assessing the determinants of the flows of collaboration between 74 305 regional dyads from the 386 NUTS3 regions that have at least one publication in the period.

We embed the 'network closure' variable in gravity equations and use a zero-inflated negative binomial to estimate the model. Also we control for three separation effects commonly used in the literature: geographical distance, spatial contiguity and national borders (e.g. [Maggioni et al., 2007](#); [Scherngell and Barber, 2009](#); [Hoekman et al., 2010](#)). We find that the 'network closure' has a positive and statistically significant effect on knowledge flows. Moreover, adding this variable to the model significantly alleviates the negative effects of geographical distance and of national borders. The attenuation of the national border effect is particularly strong, suggesting that 'network proximity' may be a favored channel for international collaborations.

The remaining of the paper is organized as follows: first the determinants of inter-regional collaborations are discussed; Section 3 then presents a methodology designed to measure regional closure at the dyadic level; in Section 4 the data set is presented as well as the empirical methodology; the empirical findings are reported and discussed in Section 5 while Section 6 concludes.

2 The determinants of inter-regional collaborations

In this section we describe the determinants of scientific collaborations. First we discuss the static ones, which depend on the characteristics of the nodes and do not evolve over time. Second we present the micro-determinants of collaboration stemming from the network. And finally, we discuss of the extension of the network determinants to the regional level.

2.1 Static determinants of collaboration

When it comes to analyze the determinants of collaboration, the concept of proximity prove to be a very useful framework ([Boschma, 2005](#)). By distinguishing several types of proximity between agents (such as geographical, institutional, cognitive or organizational), this

framework allows to analyze each of them and to easily assess their interplay. Also one can distinguish two mechanisms through which proximity, whatever the form, favors collaboration: *i*) proximity augments the probability of potential partners to meet and *ii*) it reduces the costs involved in collaboration; thus rising its expected net benefits, and the likelihood of success.

Geographical proximity can be decomposed in such a way. First, the context of collaborative production of knowledge may imply the partners to share and understand complex ideas, concepts or methods; the collaboration may then involve a certain level of tacit knowledge. Consequently face to face contacts may be important in conducting effectively the research by overcoming the problem of sharing tacit knowledge (Collins, 2001; Gertler, 2003). Also face to face contacts allows direct feedbacks, eases communication and the litigation of problems and facilitates coordination (Beaver, 2001). All these elements heightens the probability of a collaboration to succeed. Thus, geographical distance, by implying greater travel costs and lesser opportunities to meet, reduces the possibilities of face to face contacts, then heightening the costs and reducing the likelihood of success of collaboration.

Second, being closer in space enhances the likelihood of potential partners to meet. Indeed, social events where researcher meet to share ideas, such as conferences, seminars or even informal meetings, are linked to geographical distance; thus heightening the chances to find the research partner at a local scale. For instance, by analyzing data on the participants at the ERSA conference, Van Dijk and Maier (2006) show that distance to the event affect negatively the likelihood to attain it. Also, the social embeddedness of researchers has been shown to decay with geographical distance (Breschi and Lissoni, 2009), such that they will have a better knowledge of the potential partners at a closer distance.

Consequently, the effect of geographical distance should be negative. This fact has been evidenced by various recent studies, in different contexts: in the case of co-authorship in scientific publications (Frenken et al., 2009; Hoekman et al., 2010, 2009), in patents (Hoekman et al., 2009; Morescalchi et al., 2014; Maggioni et al., 2007), or in the case of cooperation among firms and research institutions within the European Framework Program (Scherngell and Barber, 2009).

Turning to institutional distance, in the context of inter-regional knowledge flows, this distance is often approximated by the effect of country borders (Hoekman et al., 2009). Indeed, several features affecting knowledge flows happen at the national level (Banchoff, 2002). For instance funding schemes are more likely to be at a national scale, thus facilitating the collaborations within country. In the same vein, workers are more mobile within than across countries, and as they may keep ties with their former partners, their social network appear to be more developed at the national level. Also norms, values and language are likely to be shared within a country, facilitating collaboration. As a consequence, as for the geographical distance, the literature shows evidence that it has a negative effect over

knowledge flows (e.g. [Hoekman et al., 2010](#); [Morescalchi et al., 2014](#)).

2.2 The role of networks in the process of collaboration

In this section we introduce the endogenous characteristics that may influence the formation of collaborations. The term endogenous here is related to the notion that the connections between nodes may be dependent on the past states of the network and on the specific position of these two nodes.

One central issue in the theoretical network literature is to understand real network topologies; how do social networks form and evolve. A classic example is [Barabási and Albert \(1999\)](#) model, based on a preferential attachment mechanism. Preferential attachment occurs when the best connected nodes are the most likely to attract new links. By introducing this key element in a network formation process, [Barabási and Albert](#) have been able to create random networks whose properties are close of those of real social networks, particularly that the distribution of links among the nodes is very skewed.

Another mechanism that may play a role in the network evolution is triadic closure, defined as the propensity of two nodes that are indirectly connected to form a link; in other words, it relates the fact that social triangles are likely to close. By opposition to dyads, triads have some advantages. By reducing the individual power, triads can help to mitigate conflicts and favors trust among the individuals ([Krackhardt, 1999](#)). Then the closure may happen for the individuals to benefit from the structure of a triad. In a recent study on the German biotechnology industry, [Ter Wal \(2013\)](#) shows that triadic closure among German researchers has been increasingly important over time, as the technological regime was changing and more trust was needed among the partners. Also, by examining the behavior of Stanford's researchers, [Dahlander and McFarland \(2013\)](#) show that having an indirect partner rises significantly the probability to collaborate.

Another feature of social networks that may influence their evolution is homophily. Homophily can be seen as a compelling feature of social networks. It can be depicted as 'the positive relationship between the similarity of two nodes in a network and the probability of a tie between them' ([McPherson et al., 2001](#), p.416). This characteristic has been analyzed by sociologists in various context, like in friendships at school or working relationships, and show that similarity among individuals is a force driving the creation of ties. As [McPherson et al. \(2001, p.429\)](#) puts it: 'Homophily characterizes network systems, and homogeneity characterizes personal networks' and science is no exception. For instance [Blau \(1974\)](#) studies the relationships among theoretical high energy physicists and show that the similarity of their specialized research interest as well as their personal characteristics are important factors determining research relationships. Via the concept of homophily, one has that connected nodes tend to be similar, then researchers that are close with respect to the network

distance are more likely to be more similar than those who are far in the network distance, thus favoring the chances of collaboration.

Also the network may be seen as a provider of informational externalities, thus being critical in determining future collaborations. Indeed, as the need to collaborate is getting more and more acute (Jones, 2009), finding the right partners is critical; but may be time consuming. And time, as Katz and Martin (1997, footnote 28) point out, is one of the most important resource for researchers, even before funding. As a consequence, the network can act as a reliable repository of information in which researchers can find their future collaborators (Gulati and Gargiulo, 1999). The role of networks can then be best viewed by analogy to optimization problems: though not giving the *global* best match, the network helps to get the *local* best match. Researchers are time constrained and are not fully rational in the sense that they do not dispose of all the required information nor of the ability to gauge all potential matches to select the best one. Then 'picking' the partner in the network vicinity may be a rational and efficient choice. Two mechanisms are at play that make the network important in finding good matches: First, network proximity unveils information on the potential partners and on their reliability, specific research interests or personal characteristics; and second, through the feature of homophily, the potential partners are likely to be good matches as they are likely to be similar.

To conclude this subsection, one can say that networks can be seen as a strong determinant of its own evolution. Moreover by the formerly developed concepts, the network distance should be a determinant of collaborations, with a negative effect. At the micro level this has been evidenced by Fafchamps et al. (2010) in the context of co-authorships in economics and by Autant-Bernard et al. (2007) in the context of collaboration choices among organizations in the 6th Framework Program of the European Union.

Next subsection discuss the implication of extending this notion to an aggregated level, that is at the scale of inter-regional flows.

2.3 Networks, from the micro level to the regional level

When analyzing inter-regional networks, one has to keep in mind they are the aggregated view of micro-economic decisions. Thus it may be difficult to consider regions simply as individual agents and to apply them the same concepts as the ones used at the micro level (Ter Wal, 2011, p.1061).

Nonetheless, it can be argued that there may exist some externalities happening at the regional level, thus suggesting the feasibility of using network analytical tools at this geographical scale. As showed in the previous subsection, networks can – and by network, indirect links – be of prime importance in finding the right partner. Then inter-regional flows can involve a form of informational spillovers. These flows may carry information

on the potential partners, then providing a repository of information at the regional level. Moreover there can be externalities occurring at the organizational level that could not be observed by focusing at the micro level; for instance in the case where the information on collaborations is known at the laboratory level, the members of the lab can benefit from it. Thus analyzing network effects at the regional level is interesting as it can show patterns that may be hidden at the strictly micro-level. But care is to be taken on the nature of the inter-regional flows, they are solely the aggregation of individual collaborations. Thus one should use special tools to take care of that singularity.

Some studies make use of network analysis technique to study the network at an aggregated level, thus assuming knowledge spillovers happen at the regional scale. For instance [Maggioni et al. \(2007\)](#) use the information on inter-regional collaborations to test whether the production of knowledge of one region is dependent on those of its network neighbors. Also [Sebestyén and Varga \(2013a,b\)](#) have developed a measure to assess the 'quality' of a region's network. This measure is based on the clustering of a region's neighbors and on the level of knowledge of direct and indirect neighbors. They applied this measure to assess the knowledge productivity of regions, in the production of patents and scientific articles. In the context of the European Framework Program [Wanzenböck et al. \(2013, 2014\)](#) apply centrality measures from the sociological literature to the regional scale to assess the central position of regions and assess its determinants. In a recent study, [Broekel and Hartog \(2013\)](#) make use of exponential random graph technique to assess the determinants of knowledge flows. By doing so, they assume regions to behave as individuals (by flattening the weighted network to an unweighted one) and they transpose micro-level concepts as preferential attachment or triadic closure.

Contrary to previous studies, we will here try to develop a measure of network proximity of regions that takes into account the nature of aggregated networks in the context of collaborations. No study, to the best of our knowledge, has assessed the effect of network closure on inter-regional knowledge flows. This is because there is a lack of analytical tools to assess it. We overcome this analytical problem by providing a dyadic index of network closure relating the strength of indirect ties between two regions.

3 A measure of 'network proximity' between regions

In this section we define an index trying to catch a simple network effect: the concept of triadic closure, but applied to the regional case. We will see that the concept needs some rethinking to be extended at the regional level.

3.1 Assessing network proximity in aggregated networks

Triadic closure as an approximation of network proximity. One simple way to deal with network effects at the regional level (or aggregated level) is to extend the concept of triadic closure to the regional level. At the micro level (i.e. individual's level) this concept states that the more two agents have common collaborators, the more likely they engage in a relation. To extend this notion to the regional level, we introduce the notion of bridging path. This concept is illustrated by figure 1. To put it simply, a bridging path is a set of two collaborations from two different regions with another one such that they involve the same collaborator (i.e. the micro-agent) in the bridging region. That is to say if we consider three regions i , j and k , and have two collaborations: $i - k$ and $j - k$, then this set of ties will form a bridging path between i and j via k only if the micro-agent in k is the same that participates to both ties. The notion of bridging path between regions is then very close to the micro concept of common neighbor, the difference being that the triad of agents are in three different regions.

The main driver of the idea is that the larger the number of bridging paths between two regions, the closer they will be in the social space as their micro-agents will be more likely to share common collaborators through the third region. Thus, the higher the propensity of the two regions to form a direct tie, thanks to a higher number of social triangle between them that can be closed.

So we will make use of the concept of bridging path to assess social proximity between regions. But because of the particularity of dealing with the regional level, that is dealing with a weighted network, we have to define how the number bridging paths can be obtained.

The most intuitive and straightforward way to cope with weighted networks is to transform it into a regular unweighted network. Then as the weighted matrix becomes the adjacency matrix, regular network measures from the social network analysis literature can be used. Nonetheless, this method is very unsatisfactory, its huge drawback being that all the critical information contained in the weights are set aside by considering each existing link as equal.

Instead we here propose a framework designed to overcome the problem of using weighted networks, from which an index of social proximity can be derived. This framework has the advantages of being based on a coming and going between the micro and the macro level within the collaboration network, under several assumptions on the behavior of the micro strata. It is presented in the next subsection.

3.2 A random matching process

We define the assumptions that lead to a *random matching process*, that is used to infer the structure of the micro network by using all the information included in the observable macro

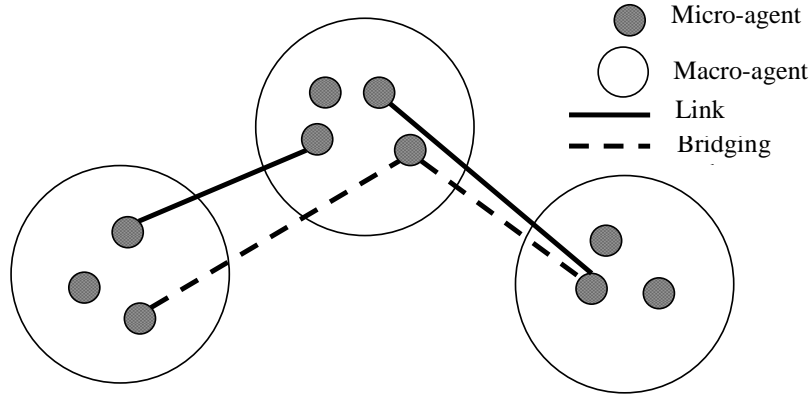


Figure 1: Illustration of a bridging path between two regions.

network. This random matching process reflects the ex post probability to be matched: that is to say the probability that two individuals for two particular regions have been matched conditional to the structure of the regional flows of collaborations. It relies on the two next assumptions.

Assumption 1. A collaboration consists of a match between two agents only.

This assumption is rather functional and is used to make the model simple and tractable. The term “agent” here is intended to be taken as a broad term: it can be either a lone researcher or a team of researchers, as teams can be fairly considered as behaving like a unique entity (Beaver, 2001; Dahlander and McFarland, 2013).

Assumption 2. Whenever a collaboration occurs between two regions, we assume that the two agents involved are matched along a random matching process. That is to say: two agents from the two regions are paired fully at random.

This is the last and most restrictive assumption that ends setting the random matching process. This assumption is used to make the model as simple as possible. And still, this assumption is also in line with the intuition as it simply states that, if we take two regions A and B , the more observed collaborations between A and B , the more likely a randomly picked agent from A has collaborated with B . This is illustrated by figure 2, where one can see that if one picks one agent from region A , it is more likely that he has collaborated with B than with C and this fact can be inferred only by looking at the aggregated network.

The very aim of this random matching process is to be simple and clear in its mechanisms. Also its purpose is not to give back the “true” micro-level network from the macro strata but instead to give a baseline for a micro-network that is likely to occur, with respect to what is observable. This matching mechanism yields the convenient possibility to get access to

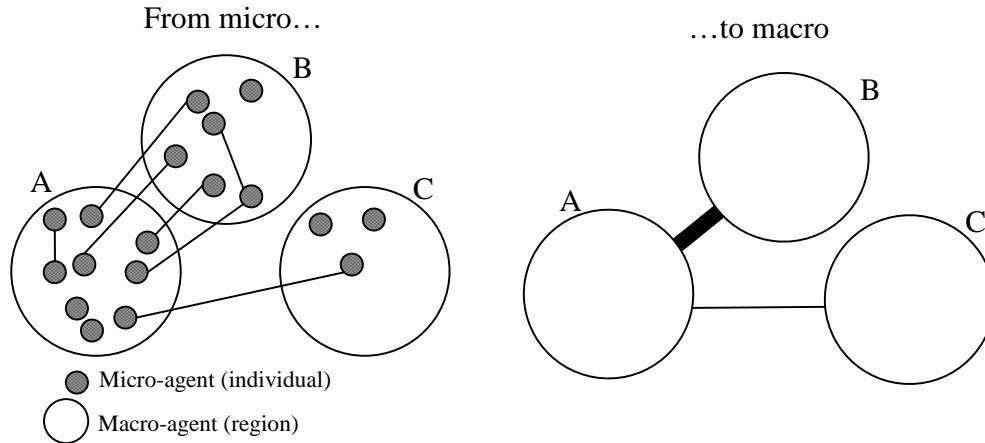


Figure 2: Illustration of the difference between the micro network of collaboration on the left leading to the observable macro network on the right.

micro-level network analysis only with the aggregated information. In particular the measure of social proximity previously defined by the bridging paths can be derived. Next subsection show how we proceed.

3.3 Building the measure

In this subsection, after some necessary notational introduction, we show how, by making use of the random matching process, we can derive the expected number of bridges between two regions.

Some notations. We consider that there are N regions. The number of collaborations¹ involving agents from regions i and j , that is the weight of the relation between i and j , is noted g_{ij} . Also g_{ii} is the number of collaborations involving only agents within region i . Let n_i be the number of agents of region i . As stated before (Assumption 1), this number is proportional to the total number of collaborations, that is formally $n_i = \alpha \sum_{j=1}^N g_{ij}$; where α is a scalar being the coefficient of proportionality between the number of collaborations and the number of agents. For simplicity we henceforth consider that $\alpha = 1$.² Henceforth, as the number of researchers of a region is directly linked to its number of publications, we make use of g_i to denote n_i .

Expected number of bridging paths (ENB). Using the information contained in the aggregated network of collaborations between regions (i.e. all the g_{ij}) along with the *random*

¹Even though collaborations and co-publications are not equivalent, we will make use of these two terms interchangeably.

²Note that setting α to 1 does not alter the reasoning as this parameter will enter the forthcoming results only as a coefficient of proportionality.

matching process assumptions previously defined, the expected number of bridging paths between two regions via another one (called the bridging region) can now be derived.

Proposition. The expected number of bridges between two regions, i and j , via the bridging region k along the random matching process is:

$$Bridges_{ij}^k = \frac{g_{ik}g_{jk}}{g_k}. \quad (1)$$

Proof. Let consider the case of three regions i , j and k , and figure out the number of bridging paths between i and j via k ; k being the bridging region. As stated before, a bridging path is a set of two collaborations: one between i and k and the other between k and j such that the researcher from k is the same in both collaborations.

Thanks to the random matching assumption, the probability that a micro-agent from k is connected to any incoming link is $1/g_k$. So if one take a given collaboration between i and k , the probability that the researcher from k is connected to a researcher from j is $1/g_k$ for any incoming link from j ; and there are g_{jk} incoming links.

Let X be the random variable that defines the number of links from j a researcher from k is connected to. As this process is a sum of g_{jk} independent Bernoulli draws of probability $1/g_k$ it then follows a binomial law: $X \sim B\left(g_{jk}, \frac{1}{g_k}\right)$ whose expected value is $\frac{g_{jk}}{g_k}$.

Each link between i and k then generates a number of bridges following the law of X . Let X_l be the random variable of the number of bridges generated by the link l between i and k . Reminding that there are g_{ik} links between i and k , we can define the number of bridging paths between i and j as the sum of all X_l , that is $Bridges_{ij}^k = \sum_{l=1}^{g_{ik}} X_l$. Then it follows that the expected number of bridges is: $E\left(Bridges_{ij}^k\right) = E\left(\sum_{l=1}^{g_{ik}} X_l\right) = \sum_{l=1}^{g_{ik}} E(X_l) = g_{ik} \times \frac{g_{jk}}{g_k}.$ \square

The result of equation 1 takes a very simple form: it depends on the links with the bridging regions and on its size. Basically the effects behind the idea of the expected number of bridging path can be split in two main characteristics:

Intensity *Ceteris paribus*, the rise of collaborations with the bridging region heightens the expected number of bridges between i and j . This is because it heightens the share of regions i or j in k 's total collaborations, thus rising the probability that an agent from k is a bridge between the two regions.

Dispersion *Ceteris paribus*, the larger the bridging region k , the lower the expected number of bridges between i and j . This is because the incoming links of i and j will be scattered in a larger number of partners from k , thus lowering the probability that a micro-agent is a bridge specifically between i and j .

So the ENB is defined, ruled by the intensity and dispersion effects. But two regions are not restricted to having only one common collaborator. Thus we have to compute the total number of bridging paths, that is the bridging paths stemming from all the network-neighboring regions.

Total expected number of bridging paths (TENB). Now that the expected number of bridges stemming from a specific region is properly defined by equation 1, we can easily define the total expected number of bridges between two regions as the sum of the bridges stemming from all their neighboring regions:

$$Bridges_{ij} = \sum_{k \neq i,j} \frac{g_{ik}g_{jk}}{g_k} \quad (2)$$

A major property of the total expected number of bridges is:

Simultaneity Regions i and j can have a high number of bridges between them if they both do collaborate intensely with the same regions.

The simultaneity effect means that the distribution of the collaborations among the partners is of prime importance; it is not sufficient to have a common neighborhood as measured by the adjacency matrix (which consider flows of one collaboration equal to flows of hundreds), regions really need to collaborate intensely and simultaneously with the same neighbors to have a high number of bridges. So this is a feature for which the weights of the network carry information of utmost importance. Without the weights of the network, there is no simultaneity, intensity, nor dispersion effect thus losing all the refinements.

The main purpose of the measure of the total expected number of bridges from the *random matching process* is to give a baseline for the underlying real number of bridges; that is to say a measure that is likely to relate a “true” fact that can be measured at the micro level. When one looks at the form of the total expected number of bridges, summarized by equation 2, one can see that in fact it relates very intuitive facts: It simply states that the more regions have common neighbor and the more they do collaborate simultaneously intensely with them, the more likely they are to share a common social space.

3.4 Illustration

We here analyze the index and the expected number of bridges by using simple examples that are common in the social networks literature, that is to say the full network, the circle network and the star network (Jackson, 2010). These examples and the values of the index associated to them are presented on figure 3. We consider that the nodes represent the regions and that each link connecting them is of value θ .

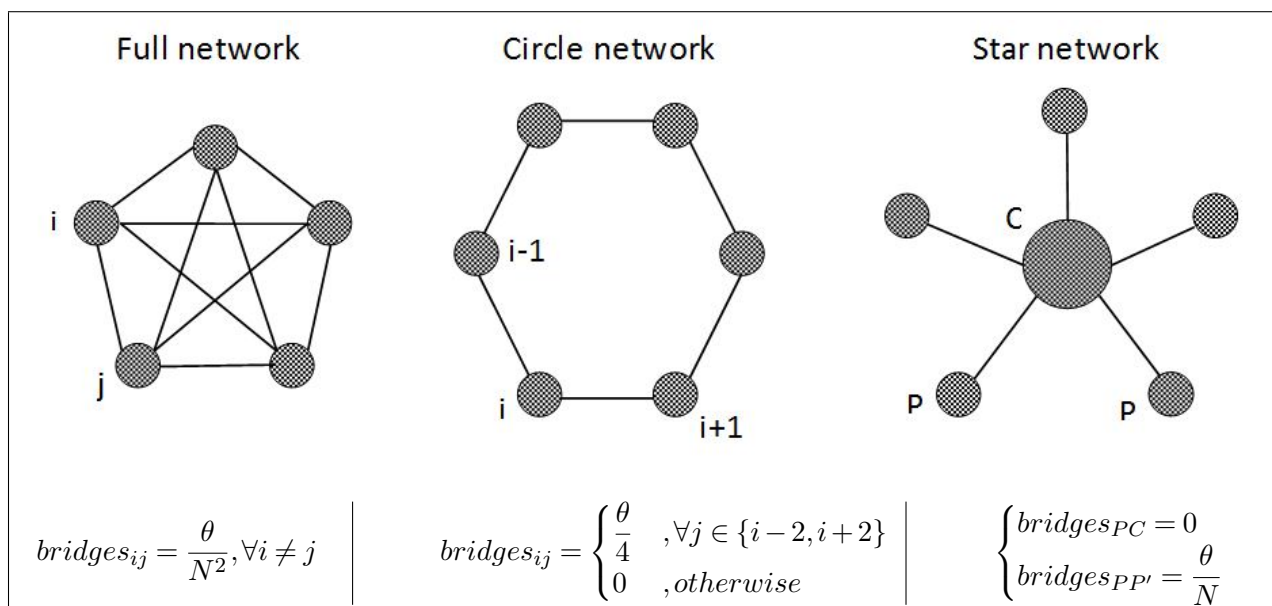


Figure 3: Example of the expected number of bridges for different basic networks. We assume that each node has a total of θ links that is evenly split across its neighbors. The drawings are done for a fixed number of nodes, but the measures associated to them are done for $N + 1$ nodes (in the star network there is one core node, noted C , and N peripheral nodes, noted P), so that the numbers between are directly comparable.

The full network. The full network is a situation where all the nodes are connected. All dyads have an index with a relatively low value, that is because the simultaneity is low (the links are evenly spread along all the nodes instead of being pooled) and the dispersion is high (the share of one link on the total links of a given is low).

The circle network. The situation depicted by the circle network is much different from the full network. The TENB is high between two nodes that are at geodesic distance 2. That is because they do collaborate intensely with the same partner with only a slight dispersion effect. On the opposite, the TENB is nil for all other dyads, even when two nodes are directly connected. That is because this measure takes into account only the common network-neighbors of the regions and not their direct link.

The star network. In the star network, there is no closure between the node of the core and the nodes of the periphery because peripheral are only tied to the core, thus the core and the periphery cannot have common network-neighbors. The TENB for two nodes of the periphery is relatively low, because despite that all are linked to the same node (high simultaneity), this is compensated by the dispersion effect that is high (the core region is large).

4 Data and methodology

In this section we first introduce the methodology, then the construction of the data set and all the variables is explained. Some descriptive statistics are given last.

4.1 Empirical model

In order to assess whether the network distance influence inter-regional collaboration flows, we will make use of the gravity model. The gravity model is a common methodological tool used when assessing spatial interactions in various contexts such as trade flows or migration flows (Roy and Thill, 2004; Anderson, 2011), and it is also be applied to knowledge flows (Maggioni et al., 2007; Autant-Bernard et al., 2007; Maggioni and Uberti, 2009; Hoekman et al., 2013). In a nutshell, the gravity model reflects the idea that economic interactions between two areas can be explained by the combinations of centripetal and centrifugal forces; while the masses of the regional entities act as attractors, the distance separating them hampers the attraction. It can be written as:

$$Attraction_{ij} = Mass_i^{\alpha_1} Mass_j^{\alpha_2} F(Distances_{ij}), \quad (3)$$

with $F(\cdot)$ being a decreasing function of the distances; for instance it can be, with x the distance, $F(x) = 1/x^\gamma$ or $F(x) = \exp(-\gamma x)$ (Roy and Thill, 2004). Traditionally, $Mass_i$ and $Mass_j$ are respectively called mass of origin and destination; also in the case where $Attraction_{ij}$ is undirected, as in collaboration networks, α_1 should be equal to α_2 . We apply this methodology to the context of inter-regional collaborations, here captured by the $Attraction_{ij}$ variable. Before giving the equation of the model to be estimated along the gravity model, we first introduce how the data set was built.

4.2 Data

To measure the intensity of collaborations among two regions, we will make use of co-publication data³. We approximate collaborations among researchers by co-publications, as in other studies (e.g. Hoekman et al., 2009; Ponds et al., 2007).

We extracted the information on co-publications from the Thomson-Reuters Web of Science database. This database contains information on papers published in most scientific journals, with, for each article, the list of all the participating authors along with their institutions.

³Publications can be seen as the result of successful collaborations and by definition they do not reflect all the collaborations occurring in a given period of time. Nonetheless, as Dahlander and McFarland (2013, p.99) puts it, along a study using extensive data from research collaborations at Stanford university, “*published papers afford a visible trail of research collaboration*”.

The data is extracted for a time period ranging from 2001 to 2005 and the geographical scale is restricted to the five largest European countries: Italy, France, Germany, Spain and the United Kingdom, as in [Maggioni et al. \(2007\)](#). Also to avoid the problems that can arise when mixing several disciplines because of researchers' behavior and publishing schemes that may differ between fields, we restrict the analysis to one specific field: Chemistry.

For each paper, this database reports the institution of each author's by-line. As there is an address assigned to each institution, it is possible to pinpoint each of them. The localization was mainly done using the postcodes available in the addresses, which should be a very reliable determinant of location. More than 85% of the sample could be assigned a location with the postcodes. The remaining 15% were located using an online map service⁴ with the information on the name of the city and the country. In the end, 99,6% of the sample was located⁵. Once located, each institution is assigned to a NUTS3 region with respect to their latitude/longitude coordinates. We discarded the NUTS3 regions that produced no article during the period, as they bring no information and only artificially rise the number of regions⁶. Thus from the initial 820 NUTS3 regions, we keep only the 386 that are active in chemistry.

To sum up, the database consists of papers from Chemistry journals, with at least one author affiliated to an institution from the selected countries, for a total of 125,075 publications distributed along 386 NUTS3 regions and 5 years.

4.3 Creating the variables

Dependent variable. First we create the dependent variable, that is $Copub_{ij}$, the number of co-publications involving authors from both regions i and j , in the time period 2004-2005. Several methods could have been used to build this variable. Mainly there are the 'full count' or the 'fractional' methodology⁷. The former gives a unitary value for each dyad participating to a publication, while the latter weights each publication by the number of participants such that the higher the number of participants, the lower the value each dyad receives; for instance if there are n participants, each dyad receives $1/n$. As in other studies ([Frenken et al., 2009](#); [Hoekman et al., 2010](#)), we make use of the full count methodology as it relates to the idea of participation to knowledge production instead of net contribution to knowledge production [OST \(2010, p.541\)](#).

⁴The online map service used was Google Maps©.

⁵Despite its simplicity, the accuracy of the location with only the name of the city and the country is quite high. Indeed, we located all addresses with both methods: city/country or postcodes. When comparing the two methods, one can see that only less than 1.5% of the NUTS3 codes differ between the two methodologies. This number falls to less than 0.4% when considering the NUTS2 codes.

⁶Also within a gravity model framework, depicted by equation (3), those regions would have a mass of zero, so that their level of attraction with any other region would be zero per construction.

⁷Using the fractional count instead of the full count methodology do not alter the results. The results-tables with fractional counting can be sent on demand.

The masses. The mass of a region is defined as the total number of papers produced from researchers whose institutional address is within the region. This mass represents a natural force of attraction. It is based on the years 2001 to 2003.

The variable of main interest: the expected number of bridges. Here we create the variable that relates to the network proximity of regions. From the model developed in section 3, we derived the closed form of the expected number of bridges between two regions along a random matching process. The network proximity is approximated by the number of bridges with the idea that the higher the expected number of bridges between two regions, the more they benefit from informational externalities from their neighbors so that it should affect positively their number of collaborations. Let $Bridges_{ij}$ be this variable whose definition is in equation (2). This variable is built using a three years rolling windows, from 2001 to 2003, such that:

$$Bridges_{ij}^{t=2001-2003} = \sum_{k \neq i, j} \frac{Copub_{ik}^{t=2001-2003} Copub_{jk}^{t=2001-2003}}{Mass_k^{t=2001-2003}}, \quad (4)$$

where $Mass_k^{t=2001-2003}$ is the total number of publications from region k ⁸. As clear in its definition, $Bridges_{ij}$ is fully independent of $Copub_{ij}$, the direct linkage between i and j , so that it depends only on the structure of their neighboring regions.

Other independent variables. To capture the impeding effect of the geographical distance, we build the variable $GeoDist_{ij}$ that is the geographical distance between two regions in kilometers measured by the as the crow flies distance between the geographic centers (centroids) of the two regions. To catch the effect of institutional distance, we use a cross-country dummy $CountryDist_{ij}$ taking the value of 1 when the two regions are from different countries and 0 otherwise.

Also to further take into account the notion of geographical proximity, we create a variable of regional contiguity, as geographical distance alone may not grasp all the effect of this notion. Indeed, for instance two regions that are geographical neighbors may be likely to some common culture, even if they are from different countries, thus mitigating the border effect. Consequently we create the variable $ContigDist_{ij}$ which is 0 when regions i and j are contiguous and value 1 otherwise. It should have the same effect as geographical distance: contiguous regions are expected to collaborate more than non contiguous ones.

⁸The denominator in equation (4) should be the number of researchers. The problem of using the number of researchers is that we do not know precisely their affiliation. Then we chose to approximate the number of researchers by the number of papers produced from this region; supposing a proportionality between these two numbers. We tried other measures for the denominator, such as the total number of researchers or weighted number of researchers; they implied no significant change to the results.

Year range of the variables. As the analysis is cross sectional, we separate the construction of the variables $Bridge_{ij}$ and $Mass_i$ from the variable $Copub_{ij}$ so that they do not occur simultaneously. The first two are based on the network of co-publication for the years 2001-2003 while $Copub_{ij}$ is built for the years 2004-2005.

4.3.1 Model and estimation procedure

Based on the gravity model and on the previously defined variables, the model we will use has the following form:

$$E(Copub_{ij}) \sim \exp\left(\beta_0 + \beta_1 [\ln(Mass_i) + \ln(Mass_j)] + \beta_2 \ln(GeoDist_{ij}) \right. \\ \left. + \beta_3 \ln(Bridges_{ij} + 1) + \beta_4 ContigDist_{ij} + \beta_5 CountryDist_{ij} \right) \quad (5)$$

There is a need of adding 1 to the variable $Bridges$ as its value is nil for all the regional dyads that have no common collaborator. As $Copub_{ij}$ is by nature in the form of count data, a natural way to estimate equation (5) is via a Poisson regression. Nonetheless, Poisson regressions are constrained by the fact that the variance is by definition equal to the mean. The negative binomial (NB) is a less constrained model that allows for over-dispersion of the data by adding a parameter that modifies the variance. Also as there are large disparities among regional pairs, over-dispersion may be at work (see for instance the descriptive statistics of next subsection).

Also, as the focus is on bilateral relations, the likelihood of the absence of collaboration between two regions is high as the matrix of collaboration is very sparse. That is to say, there is a high number of cases where the dependent variable is nil. Thus this suggests that there may be first a process that determines the probability of the dependent variable to be zero or non-zero. Such a model is the zero-inflated negative binomial (ZINB) that will be used to conduct this analysis, as in other studies on inter-regional knowledge flows (e.g. [Hoekman et al., 2009](#); [Frenken et al., 2009](#); [Morescalchi et al., 2014](#)).

The ZINB is split in two processes. The first process, the zero inflation (ZI), determines the probability of being zero along a Logit density. The second process is the NB and gives the number of counts. Let y_{ij} be the dependent variable. Thus, knowing that $y_{ij} = 0$, the zero can be generated either by the Logit process or by the NB process:

$$L(y_{ij}|y_{ij} = 0, \mu_{ij}^{NB}, \mu_{ij}^{ZI}, \theta) = \underbrace{\frac{\exp(\mu_{ij}^{ZI})}{1 + \exp(\mu_{ij}^{ZI})}}_{0 \text{ from the Logit}} + \underbrace{\frac{1}{1 + \exp(\mu_{ij}^{ZI})}}_{\text{not 0 from the Logit}} \times \underbrace{\left(\frac{\theta}{\theta + \exp(\mu_{ij}^{NB})} \right)^\theta}_{0 \text{ from the NB}}$$

The likelihood of the values that are strictly positive has the following form:

$$L(y_{ij}|y_{ij} > 0, \mu_{ij}^{NB}, \mu_{ij}^{ZI}, \theta) = \underbrace{\left(\frac{1}{1 + \exp(\mu_{ij}^{ZI})} \right)}_{\text{not 0 from Logit}} \underbrace{\frac{\Gamma(\theta + y_{ij})}{\Gamma(\theta)\Gamma(y_{ij})} \left(\frac{\theta}{\theta + \exp(\mu_{ij}^{NB})} \right)^\theta \left(\frac{\exp(\mu_{ij}^{NB})}{\theta + \exp(\mu_{ij}^{NB})} \right)^{y_{ij}}}_{\text{the NB}}.$$

The first term in parenthesis is the probability of not being zero along the ZI process, the other elements on the right side are the likelihood of the NB process. The parameter θ is called the shape parameter and represents over-dispersion. The variance of the NB is $\mu^{NB} + (\mu^{NB})^2/\theta$, so that over-dispersion increases as θ decreases, and conversely. Also, at the limit, when θ tends to infinity the negative binomial density resumes to a Poisson density.

Both processes of the ZINB depend on a set of parameters, here μ_{ij}^{ZI} and μ_{ij}^{NB} . Also the factors that may influence the numbers of counts should also influence the ZI process, so that the same explanatory variables are kept for the two processes, as defined by equation (5). So with $m \in [ZI, NB]$, we have:

$$\begin{aligned} \mu_{ij}^m &= \beta_0^m + \beta_1^m [\ln(Mass_i) + \ln(Mass_j)] + \beta_2^m \ln(GeoDist_{ij}) \\ &\quad + \beta_3^m \ln(Bridges_{ij}) + \beta_4^m ContigDist_{ij} + \beta_5^m CountryDist_{ij} \end{aligned} \quad (6)$$

this equation defines the model that will be estimated in the ZINB regression.

4.4 Descriptive statistics

The data set is composed of all the bilateral relations among 386 NUTS3 regions, which leads to 74 305 ($= 386 \times 385/2$) observations or regional pairs. Table 1 shows some descriptive statistics on the data set and the main constructs. Looking at the number of collaborations, one can see that the distribution is very uneven, with a coefficient of variation of 7.5, giving some vindication for the use of the NB. On top of that, the fraction of inter-regional relations that are nil is very high, with more than 93% of zeros, thus the use of a zero inflated process seems justified.

The variable approximating the network proximity, that is the expected number of bridges defined by equation (4), is also unevenly distributed, but less than the number of co-publications, with a coefficient of variation of 4.23. Also there is much less zeros than for collaborations as this measure depends on the common network-neighbors and has a non-negative value whenever two regions share a common neighbor. Also, more than 47% of the regional pairs have a common neighbor (i.e. both regions have at least one collaboration with a third common region).

Table 1: Descriptive statistics of the main variables of the collaboration network.

	Min	Max	Mean	SD	CV	Freq. 0
Inter-regional collaborations	0.00	123.00	0.18	1.35	7.50	93.8%
Expected number of bridges (network proximity)	0.00	29.01	0.13	0.55	4.23	52.1%
Distance in kilometers	5.43	2760.65	993.40	538.33	0.54	-
Total publications, region i (Mass origin)	1	1231.34	97.32	160.59	1.65	-
Total publications, region j (Mass destination)	1	1231.34	92.58	153.22	1.65	-

Table 2: Correlation matrix of the covariates.

	1	2	3	4	5	6
1 $Mass_i(\ln)$	1.00					
2 $Mass_j(\ln)$	-0.00	1.00				
3 $Bridges_{ij}(\ln)$	0.38 *	0.38 *	1.00			
4 $CountryDist_{ij}$	0.01	0.06 *	-0.23 *	1.00		
5 $ContigDist_{ij}$	-0.01	0.01	-0.06 *	0.16 *	1.00	
6 $GeoDist_{ij}(\ln)$	0.01 *	0.05 *	-0.18 *	0.71 *	0.30 *	1.00

*: statistically significant at the 1% level (Pearson correlation).

Contrary to the two previous variables, which are at the dyadic level, the distribution of the masses are much more evenly distributed, as they reflect information at the node level.

Table 2 relates the correlation among the main variables. As expected, the highest correlation is between the geographical distance and the country border variable. From the correlation matrix, there is no sign of a multi-collinearity problem.

5 Results

The results of the zero-inflated negative binomial regression are shown in table 3. This table reports the results of the ZINB, split in two parts: the ZI part relates to the probability of being zero, thus positive estimates in this part indicates a negative effect on the dependent variable as it increases its probability of being zero. The second part is the NB that estimates the counts where a positive estimate imply a positive effect on the dependent variable.

The estimation is done to assess network proximity as a determinant of inter-regional knowledge flows, and see its interplay with other classic explanatory variables. Thus, in model 1, equation (6) is estimated without the variable $Bridges_{ij}$. This variable is then included in the second model.

Consistent with the literature (Hoekman et al., 2009, 2010; Scherngell and Barber, 2009), the effects of the geographical is negative, as it is the case for the institutional distance (measured by the $CountryDist$ variable) and for the non-contiguity.

As there are two parts in the ZINB, the interpretation of the coefficients is not straightfor-

Table 3: Zero-inflated negative binomial regression model of inter-regional collaborations in chemistry for the years 2004-2005. The independent variable *Bridges* is a proxy for network proximity and is built using information on the collaboration network for the years 2001-2003.

Dep. Var.: the number of collaborations	model 1	model 2
<i>Negative Binomial part</i>		
Constant	-1.8868* (0.1949)	0.1631 (0.1897)
Mass origin / destination (ln)	0.5781* (0.0154)	0.2098* (0.0195)
Geographical Distance (ln)	-0.2871* (0.0308)	-0.2111* (0.0284)
Non-contiguity	-1.1331* (0.1019)	-1.0429* (0.0884)
Different country [institutional distance]	-1.0547* (0.0583)	-0.5248* (0.059)
Bridges (ln) [network proximity]		1.0582* (0.0486)
<i>Zero-Inflated part (explain prob. being zero)</i>		
Constant	2.8599* (0.3994)	3.0377* (0.3451)
Mass origin / destination (ln)	-0.7332* (0.0233)	-0.6044* (0.03)
Geographical Distance (ln)	0.2434* (0.0634)	0.2299* (0.0619)
Non-contiguity	1.0574* (0.2834)	1.0093* (0.2159)
Different country [institutional distance]	1.9131* (0.1207)	1.4292* (0.1233)
Bridges (ln) [network proximity]		-2.7506* (0.228)
<i>Fit statistics</i>		
Over-dispersion (θ)	1.0806* (0.0504)	1.3378* (0.0521)
Number of observations	74 305	74 305
Non-zero	4 554	4 554
Adj-pseudo R^2	0.34539	0.36758
AIC	33 060.272	31 939.309
BIC	32 977.329	31 837.933
<i>Vuong</i> $\sim N(0, 1)$	8.44*	16.21*

*: statistically significant at the 1% level. Robust standard error in parenthesis. On the concern of multi-collinearity, the highest variance inflation factor (VIF) does not exceed 2.15. The Vuong test compares the ZINB with the NB; the null hypothesis is that the two models are not distinguishable.

ward, but one can still interpret the effect of the estimates in each process. *Ceteris paribus*, if a regional dyad is across- instead of within-country, the likelihood of being zero is 6.75 times higher (i.e. $\exp(1.9131)$, in the ZI part); and the number of collaboration (if non zero) falls of 65% (i.e. $1 - \exp(-1.0547)$ in the NB part). The effect of the geographical distance can be interpreted as elasticities; the rise of the geographical distance of 1% rises the probability of being zero of 0.24% and lowers the counts of 0.28%; if the distance is doubled, the respective magnitudes are 17% and 18%. Though the effect of the institutional distance is of very high, the order of magnitude of these effects is in line with other estimates in the literature (Maggioni et al., 2007; Hoekman et al., 2009).

Also, *ceteris paribus*, being non-contiguous instead of contiguous rises the probability of being zero of 2.71 times and reduces the expected number of collaboration by 67%, so that it exhibits a neat interplay between the regions and their close geographical vicinity; it can be interpreted as the sign of localized knowledge spillovers (Maggioni et al., 2007).

In the model 2, the network proximity is introduced. Network proximity is approximated by the TENB defined in section 3 and it relates the strength of the indirect connection between a regional pair. First of all, adding this explanatory variable improves significantly the fit statistics while reducing the over-dispersion as θ gets higher. Also its effect is positive and statistically significant. In magnitude, for an increase of 1% of the bridges, the probability of being zero decreases of 2.7% and the number of collaborations increases of 1.05%.

The introduction of the network proximity had several effects on the other estimates. First there is a sharp decrease of the importance of the institutional distance, whose coefficients are much closer to 0 in both the ZI and the NB parts. The magnitude of the fall in the coefficients translates into a reduced effect of the institutional distance: now if a regional dyad is across- instead of within-country, the likelihood of being zero is 4.17 times higher and the expected number of collaboration falls of 40%; these effects are much lower than in model 1. The other distance impacted is the geographical one, which has lower estimated coefficients. On the contrary, there is no significant effect on the non-contiguity.

One interpretation is that the network proximity acts as a substitute of other forms of proximity such as institutional or geographical. It seems that there are spillovers stemming from the inter-regional network that yields regional dyads that have a high 'network proximity' to collaborate more. This effect of network proximity then in part overcomes the negative effect of different national borders, and to a lesser extent, the negative effect of geographical distance. Though, it does not suppress these distances, as they are still playing a role, only their size has been impacted. Also the institutional distance has been strongly reduced, maybe suggesting that the 'network proximity' is a channel of prime importance for international collaborations.

6 Conclusion

This paper has investigated the role of networks in the case of inter-regional knowledge flows. We first have theoretically defined a measure of 'regional closure' for the aggregated network at the dyadic level. As there are theoretical reasons to assume that inter-regional linkages may provide externalities and by then influence the network formation (Maggioni and Uberti, 2011); but methodological tools are scarce to assess these effects in the inter-regional context. Such a measure was then necessary to cope with the network determinants of inter-regional knowledge flows.

Using co-publication data on chemistry among 386 NUTS3 regions, we then estimate gravity equations using a zero-inflated negative binomial regression. Consistent with the existing literature, we find a negative effect of the geographical distance and of the national borders. We also find a positive and statistically significant effect of the 'network proximity', implying that the structure of the network in which the regional dyads are embedded matter in determining their collaborations. We interpret this positive effect as externalities of information happening at the level of regions, carried through the collaborations at the micro-level.

Interestingly, the introduction of the 'network proximity' variable strongly reduces the negative effect of country borders. This suggests that 'regional closure' may be a channel of prime importance in international collaborations. Thus when not taking into account the effect of 'network proximity', the national border effect may be much overestimated.

Linking to the ERA, even though the inter-regional network of co-publication seems to have a form of auto-organization, as the regions that are close in the network space tend to have collaborate more, the effect of geographical distance or of the border effect are still significant, so that there is still room for trying to reduce the barriers to collaborate.

References

- Aghion, P., Howitt, P., March 1992. A model of growth through creative destruction. *Econometrica* 60 (2), 323–351. 1
- Anderson, J. E., 2011. The gravity model. *Annual Review of Economics* 3 (1), 133–160. 4.1
- Autant-Bernard, C., Billand, P., Frachisse, D., Massard, N., 2007. Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies. *Papers in Regional Science* 86 (3), 495–519. 2.2, 4.1
- Banchoff, T., 2002. Institutions, inertia and European Union research policy. *Journal of Common Market Studies* 40 (1), 1–21. 2.1

- Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286 (5439), 509–512. 2.2
- Beaver, D. d., 2001. Reflections on scientific collaboration (and its study): past, present, and future. *Scientometrics* 52 (3), 365–377. 2.1, 3.2
- Blau, J. R., September 1974. Patterns of communication among theoretical high energy physicists. *Sociometry* 37 (3), 391–406. 2.2
- Boschma, R., 2005. Proximity and innovation: A critical assessment. *Regional Studies* 39 (1), 61 – 74. 2.1
- Breschi, S., Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 9 (4), 439 – 468. 2.1
- Broekel, T., Hartog, M., 2013. Determinants of cross-regional R&D collaboration networks: an application of exponential random graph models. Unpublished. 2.3
- Collins, H. M., 2001. Tacit knowledge, trust and the Q of sapphire. *Social studies of science* 31 (1), 71–85. 2.1
- Dahlander, L., McFarland, D. A., 2013. Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly* 58 (1), 69 – 110. 2.2, 3.2, 3
- European commission, 2012. Communication from the commission to the european parliament, the council, the European economic and social committee and the committee of the regions: A reinforced European research area partnership for excellence and growth. Tech. rep., COM(2012) 392 final, Brussels. 1
- Fafchamps, M., Van der Leij, M. J., Goyal, S., 2010. Matching and network effects. *Journal of the European Economic Association* 8 (1), 203 – 231. 2.2
- Frenken, K., Hoekman, J., Kok, S., Ponds, R., van Oort, F., van Vliet, J., 2009. Death of distance in science? a gravity approach to research collaboration. In: *Innovation networks*. Springer Berlin Heidelberg, pp. 43–57. 1, 2.1, 4.3, 4.3.1
- Gertler, M. S., 2003. Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography* 3 (1), 75–99. 2.1
- Gulati, R., Gargiulo, M., 1999. Where do interorganizational networks come from?. *American Journal of Sociology* 104 (5), 1439 – 1493. 2.2

- Hoekman, J., Frenken, K., Tijssen, R. J., 2010. Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy* 39 (5), 662 – 673. 1, 2.1, 4.3, 5
- Hoekman, J., Frenken, K., van Oort, F., 2009. The geography of collaborative knowledge production in Europe. *Annals of Regional Science* 43 (3), 721 – 738. 1, 2.1, 4.2, 4.3.1, 5
- Hoekman, J., Scherngell, T., Frenken, K., Tijssen, R., 2013. Acquisition of European research funds and its effect on international scientific collaboration. *Journal of Economic Geography* 13 (1), 23–52. 4.1
- Jackson, M. O., 2010. *Social and economic networks*. Princeton University Press. 3.4
- Jones, B. F., 2009. The burden of knowledge and the "death of the renaissance man": is innovation getting harder? *The Review of Economic Studies* 76 (1), 283–317. 2.2
- Jones, B. F., Wuchty, S., Uzzi, B., 2008. Multi-university research teams: shifting impact, geography, and stratification in science. *science* 322 (5905), 1259–1262. 1
- Katz, J. S., Martin, B. R., 1997. What is research collaboration?. *Research Policy* 26 (1), 1–18. 2.2
- Krackhardt, D., 1999. The ties that torture: Simmelian tie analysis in organizations. *Research in the Sociology of Organizations* 16 (1), 183–210. 2.2
- LeSage, J. P., Fischer, M., Scherngell, T., 2007. Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction model with spatial effects. *Papers in Regional Science* 86 (3), 393 – 421. 1
- Maggioni, M. A., Nosvelli, M., Uberti, T. E., 2007. Space versus networks in the geography of innovation: A European analysis. *Papers in Regional Science* 86 (3), 471 – 493. 1, 2.1, 2.3, 4.1, 4.2, 5
- Maggioni, M. A., Uberti, T. E., 2009. Knowledge networks across Europe: which distance matters?. *Annals of Regional Science* 43 (3), 691 – 720. 4.1
- Maggioni, M. A., Uberti, T. E., 2011. Networks and geography in the economics of knowledge flows. *Quality & Quantity* 45, 1031 – 1051. 1, 6
- McPherson, M., Smith-Lovin, L., Cook, J. M., 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 27, 415–444. 2.2
- Morescalchi, A., Pammolli, F., Penner, O., Petersen, A. M., Riccaboni, M., 2014. The evolution of networks of innovators within and across borders: Evidence from patent data. *IMT LUCCA EIC WORKING PAPER SERIES*. 2.1, 4.3.1

- OST, 2010. Indicateurs de sciences et de technologies. 4.3
- Ponds, R., Van Oort, F., Frenken, K., 2007. The geographical and institutional proximity of research collaboration. *Papers in Regional Science* 86 (3), 423 – 443. 4.2
- Roy, J. R., Thill, J.-C., 2004. Spatial interaction modelling. *Papers in Regional Science* 83 (1), 339–361. 4.1, 4.1
- Scherngell, T., Barber, M. J., 2009. Spatial interaction modelling of cross-region R&D collaborations: empirical evidence from the 5th EU framework programme. *Papers in Regional Science* 88 (3), 531 – 546. 1, 2.1, 5
- Sebestyén, T., Varga, A., 2013a. A novel comprehensive index of network position and node characteristics in knowledge networks: Ego network quality. In: Scherngell, T. (Ed.), *The Geography of Networks and R&D Collaborations*. *Advances in Spatial Science*. Springer International Publishing, pp. 71–97. 2.3
- Sebestyén, T., Varga, A., 2013b. Research productivity and the quality of interregional knowledge networks. *The Annals of Regional Science* 51 (1), 155–189. 2.3
- Ter Wal, A. L. J., 2011. Networks and geography in the economics of knowledge flows: a commentary. *Quality & Quantity* 45, 1059–1063. 2.3
- Ter Wal, A. L. J., 2013. The dynamics of the inventor network in German biotechnology: geographic proximity versus triadic closure. *Journal of Economic Geography*, In press. 2.2
- Van Dijk, J., Maier, G., 2006. ERSa conference participation: does location matter?. *Papers in Regional Science* 85 (4), 483 – 504. 2.1
- Wanzenböck, I., Scherngell, T., Brenner, T., 2013. Embeddedness of regions in European knowledge networks: a comparative analysis of inter-regional R&D collaborations, co-patents and co-publications. *The Annals of Regional Science*, 1–32. 2.3
- Wanzenböck, I., Scherngell, T., Lata, R., 2014. Embeddedness of European regions in European Union-funded research and development (R&D) networks: A spatial econometric perspective. *Regional Studies* (ahead-of-print), 1–21. 2.3
- Wuchty, S., Jones, B. F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. *Science* 316 (5827), 1036–1039. 1